CRYPTIC SEQUENCE SIMPLICITY, NUCLEOTIDE COMPOSITION BIAS, AND MOLECULAR COEVOLUTION IN THE LARGE SUBUNIT OF RIBOSOMAL DNA IN PLANTS: IMPLICATIONS FOR PHYLOGENETIC ANALYSES[1]

C. J. Bult,[2] J. A. Sweere,[3] and E. A. Zimmer[3]

ABSTRACT

Sequences of the large subunit (LSU) of ribosomal DNA from *Arabidopsis thaliana, Brassica napus, Sinapsis alba, Oryza sativa, Fragaria ×ananassa, Lycopersicon esculentum*, and *Citrus limon* were analyzed for nucleotide composition, presence of "cryptic" sequence simplicity, and evidence of molecular coevolution among the 12 expansion segments. The median value for GC content across the seven plants for the LSU was 56%, but the distribution of GC was nonrandom. Expansion segments were decidedly more GC rich (65% on average) than were the conserved core regions (52% on average). Only *Oryza sativa* had significant cryptic sequence simplicity, which was found to be greatest in expansion segments D8 and D12. Sequence similarity between expansion segments also was strongest in rice as determined by visual inspection of dot plots. The complex nature of sequence variation in the LSU of rDNA complicates the use of this molecule as a molecular systematic marker.

Choosing an appropriate gene(s) is one of the most important steps for any molecular systematics study. The nature of sequence variation in a given gene influences all subsequent steps in the analysis, including the quality of the multiple sequence alignment (which serves as the basis for hypotheses of positional homology for each nucleotide or amino acid), and the outcome of the algorithm used for phylogenetic reconstruction. Friedlander et al. (1992) described several criteria for selecting "ideal" molecular systematic markers, including (1) genes that are present in single copies (or if present in multiple copies can be differentiated one from another or are homogeneous within a species), (2) genes or regions that are longer than 500 base pairs, (3) genes that contain both conservative and variable regions, (4) genes that lack significant nucleotide composition bias, and (5) genes that lack many or long introns. In addition to these structural/compositional factors, it also is desirable to know something of the modes of sequence variation patterns across a molecule to assess the potential for non-independence of characters and other violations of the assumptions which are made by different algorithms used for phylogenetic reconstruction.

The molecular biology, evolution, and general biosystematic utility of rDNA have been reviewed extensively and will not be addressed in detail here (Arnheim, 1983; Dover & Flavell, 1984; Gerbi, 1985; Appels & Honeycutt, 1986; Flavell, 1986; Schaal & Learn, 1988; Jorgensen & Cluster, 1988; Knaack et al., 1990; Larson, 1991; Hillis & Dixon, 1991). For systematic analyses, the small subunit of rDNA has been used most often, especially for deep evolutionary divergences (e.g., Zimmer et al., 1989; Nickrent & Soltis, 1995, this issue). Recently, the internal transcribed spacer (ITS) regions that flank the 5S coding region in the rDNA repeat have gained attention as suitable markers for inter- and intra-generic comparisons in plants (e.g., Baldwin, 1992; Baldwin et al., 1995, this issue). The LSU has been used to a limited extent in plant systematics and all of the published nucleotide se-

[1] The authors are grateful to John Hancock (Australian National University) and Gabriel Dover (Cambridge University) for providing the coordinates for expansion segments in rice, and to Paul S. Hinkle for his assistance with the graphics.

[2] Molecular Systematics Laboratory, The Institute for Genomic Research, 932 Clopper Road, Gaithersburg, Maryland 20878, U.S.A.

[3] Laboratory of Molecular Systematics, National Museum of Natural History, Smithsonian Institution, Washington, D.C. 20560, U.S.A.

TABLE 1. Complete nuclear LSU rDNA sequences for plants in GenBank.

Taxon	Common name	GenBank accession	Reference
Arabidopsis thaliana	mouse-ear cress	X52320	Unfried & Gruendler, 1990
Brassica napus	rapeseed	D10840	Okumura & Shimada, 1992
Sinapsis alba	mustard	X57137	unpublished
Citrus limon	citrus	X05910	Kolosha & Fodor, 1990
Fragaria ×ananassa	strawberry	X58118	unpublished
Lycopersicon esculentum	tomato	X13557	Kiss et al., 1989
Oryza sativa	rice	M11585	Sugiura et al., 1985

quence-based analyses to date have used partial LSU sequence data (Hamby & Zimmer, 1988; Kantz et al., 1990; Zechman et al., 1990; Bult & Zimmer, 1993).

The rDNA LSU is a mosaic consisting of cores of highly conserved regions interspersed with 12 regions of variable size called "expansion segments" (Clark et al., 1984) or "divergent domains" (Hassouna et al., 1984). The number and relative positions of the expansion segments within the LSU are highly conserved among a wide range of taxa. Most of the fluctuations in length associated with expansion segments reflect the gain and loss of short, directly repetitive sequence motifs by a slippage-like mechanism during turnover (Hancock & Dover, 1988). Tautz et al. (1986, 1988) coined the term "cryptic sequence simplicity" to describe the sequence footprints of the motifs after they are shuffled among themselves by repeated slippage events. Unlike tandem arrays of unshuffled direct repeats (pure simplicity), cryptic simplicity is not easy to detect by simple visual inspection of a sequence. The elevated rates of sequence variation observed in expansion segments suggest that these regions lack function and, therefore, tolerate relatively high rates of point mutations and structural changes (Gerbi, 1985; Hancock & Dover, 1988; Tautz et al., 1988). However, the following three lines of evidence suggest that expansion segments do play a functional role, or at least evolve under selective constraints: (i) transcripts of the expansion segments are present in the mature 28S rRNA of some eukaryotes (Hassouna et al., 1984), (ii) dot plot comparisons of the complete LSU sequence from *Drosophila* demonstrate that expansion segments coevolve (Hancock & Dover, 1988), and (iii) expansion segments show a high degree of secondary structure conservation (Hancock et al., 1988). One hypothesis for the observed patterns of sequence change among and within the expansion segments is that they play some role in maintaining the steric integrity of the ribosome (Gerbi, 1985).

The complex mode of sequence variation observed in the LSU potentially complicates the use of this molecule as a phylogenetic marker. Motif shuffling and molecular coevolution among expansion segments will violate assumptions of character independence and compromise hypotheses of homology at many nucleotide positions in a multiple-sequence alignment. The purpose of the present study is to report on the nature of sequence variation across the large subunit (LSU) of ribosomal DNA (rDNA) in plants and discuss the implications of the observed patterns of variation for using the LSU as a molecular marker in phylogenetic analyses. We compare the conserved core and expansion segment region for full-length sequences of the large subunit of rDNA in plants to (1) assess nucleotide composition, (2) test for cryptic sequence simplicity, and (3) look for evidence of molecular coevolution among expansion segments.

MATERIALS AND METHODS

FULL-LENGTH LSU SEQUENCES FROM GENBANK

Seven full-length rDNA LSU sequences were found in version 79 of GenBank. The names and database accession numbers of these sequences are given in Table 1. Many more partial LSU sequence fragments are available in GenBank, but were not included in this analysis.

SEQUENCE SIMPLICITY

Sequence simplicity was analyzed using the SIMPLE34 program (Tautz et al., 1986; Hancock & Armstrong, 1994) running on a Sun SPARCstation 2 computer. The SIMPLE34 program is available via anonymous ftp or gopher from life.anu.edu.au or by sending an electronic mail message to John.Hancock@anu.edu.au. Sequence simplicity is evaluated by counting the number of tri- and tetranucleotides along the length of a DNA sequence using a 32-bp sliding window. The minimum length of a sequence that can be analyzed

by this method is 68 bp (Hancock & Armstrong, 1994). The statistical significance of the trimer and tetramer count is determined by comparing the results for the test sequence to values obtained for a set of ten randomly shuffled sequences (10,000 nucleotides long) having the same base composition as the test sequence. The score comparing the test sequence to the random sequence is called a Relative Simplicity Factor (RSF). An RSF value greater than 1 for a test sequence is statistically significant if the raw simplicity factor score for the test sequence is greater than 3 standard deviations of the simplicity factor value for the ten randomized sequences. RSF values significantly greater than 1 have cryptically simple direct repeats; RSF values significantly less than 1 have cryptically simple inverted repeats.

An underlying bias in dinucleotide composition for a particular sequence could, in turn, bias the statistical assessment of trimer and tetramer motifs in the RSF analysis. To correct for underlying dinucleotide composition bias, sequences that gave significant RSF scores were re-analyzed using a 2nd order Markov rule (Hancock & Armstrong, 1994).

DOT PLOT ANALYSIS

Dot plot comparisons were performed using the ABI Inherit Analysis software (v. 1.1; Applied Biosystems, Inc.) on a MacIntosh Quadra 900 computer to examine levels of sequence similarity between expansion segments (internal sequence similarity). The LSU of *Escherichia coli* (Brosius et al., 1980; GenBank Accession V00331) was included as a negative control as it contains no expansion segments (Clark et al., 1984; Hancock & Dover, 1988); the LSU of *Homo sapiens* (Gonzalez et al., 1985; GenBank Accession M11167) was included for comparison because it shows strong patterns of expansion segment similarities (Hancock & Dover, 1988). Internal sequence similarity was scored as present or absent by visual inspection of the dot plots. Dot plots were generated using a sliding window of 9 bp (with an offset of 3 bp) and an error tolerance of 33%. These parameters gave the best signal-to-noise ratio for detecting evidence of internal sequence similarity in plants. Because the human sequence has stronger patterns of internal sequence similarity than that observed for plants, the window was expanded to 10 bp (offset of 3; error tolerance of 20%) to decrease the background. Sequence simplicity profiles generated by the SIMPLE34 program were plotted along the X and Y axes of each dot plot to aid in visual interpretation. The profiles are graphical representations of levels of sequence simplicity over the length of the test sequence, averaged over blocks of 10 nucleotides.

TERMINOLOGY

We use the term expansion segment instead of divergent domain in this report because of the multiple meanings of the word *domain* in molecular biology. However, we retain the standard naming scheme of D1 through D12 to refer to individual expansion segments.

Results and Discussion

NUCLEOTIDE COMPOSITION

The average GC content of plant LSU sequences is 56% (Table 2). The distribution of GC is not, however, randomly distributed throughout the length of the sequence. For all seven species examined, expansion segments have a higher average GC content (65%) than do the conserved core regions (52%).

SEQUENCE SIMPLICITY

Of the seven plant species analyzed, only rice (*Oryza sativa*) has statistically significant sequence simplicity (Table 2). The significance is retained even after a 2nd order Markov correction for underlying dinucleotide composition bias (data not shown). Although their overall RSF values are not significant, three of the taxa examined, *Arabidopsis*, *Sinapsis*, and *Fragaria*, show over-representation of a particular tetramer motif (CGGC, CGGA, and AAAG, respectively; data not shown). Analysis of the 12 expansion segments individually (Table 3) reveals that much of the overall simplicity in plants is concentrated in expansion segments D8 and D12. The RSF values for D8 are significant for *Arabidopsis*, *Citrus*, *Fragaria*, *Lycopersicon*, and *Oryza*; the RSF values for D12 are significant for *Arabidopsis*, *Citrus*, and *Oryza*. After correcting for underlying dinucleotide composition bias, only the RSF values for *Oryza* remain statistically significant. In some vertebrates (human, mouse, rat), RSF values are greatest in expansion segments D2, D6, D8, and D12 (Hancock & Dover, 1988).

We also applied sequence simplicity analysis to published full-length sequences of the LSU from plant chloroplast and mitochondrial genomes and to nuclear small subunit (16/18S, SSU) rDNA. The RSF values for these sequences are not statistically significant (data not shown).

TABLE 2. Nucleotide composition, relative simplicity factor (RSF) values, and internal sequence similarity for full-length sequences of LSU rDNA in plants. *Homo sapiens* and *Escherichia coli* are included for comparison (see text).

Taxon	Length (bp)	% GC*			RSF	Internal sequence similarity‡
		ALL	CORE	ES		
Arabidopsis thaliana	3375	55	52	63	1.031	+/−
Brassica napus	3378	54	51	61	0.984	−
Sinapsis alba	3381	55	52	61	1.015	−
Citrus limon	3393	57	53	68	1.080	+
Fragaria ×ananassa	3377	56	52	66	1.039	+/−
Lycopersicon esculentum	3381	57	52	66	1.063	−
Oryza sativa	3377	59	53	71	1.184†	+
Homo sapiens	5025	69	nd	nd	1.681†	++
Escherichia coli	2904	53	nd	nd	1.015	−

* ALL = full-length sequence; CORE = conserved core regions; ES = expansion segments; nd = not determined.
† Denotes statistically significant RSF.
‡ Absence (−) or presence (+) of internal sequence similarity is determined by visual inspection of dot plots (Fig. 2).

The positions of the 12 expansion segments (D1–D12) relative to the sequence simplicity profile peaks for the LSU rDNA of rice are shown in Figure 1. The expansion segment coordinates of rice (unpublished data provided by John Hancock and Gabriel Dover) were mapped to a multiple-sequence alignment of the LSU rDNA from all seven plant species to determine position and length variation in the expansion segments and the conserved core regions. Relative positions of expansion segments and the core regions are conserved across all of the plant species analyzed. Some length variation was observed between species for the expansion segments and the conserved core regions. Most of the length variation in the conserved core regions was due to single position insertion/deletion events. The largest insertion/deletion event observed in any conserved core was a single gap involving 4 contiguous positions in the alignment. In contrast, most of the length variation in the expansion segments was due to insertion/deletion events involving at least 2 contiguous positions in the multiple-sequence alignment. The largest insertion/deletion event was observed in expansion segment D2 and

TABLE 3. Relative simplicity factor values for expansion segments (longer than 68-bp) of plant LSU of rDNA. Nucleotide composition (in %GC) given in parentheses.

Taxon	Expansion segment [median size in bp]												
	D1 [150]	D2 [224]	D3 [107]	D4 [8]	D5 [39]	D6 [27]	D7a [41]	D7b [27]	D8 [141]	D9 [24]	D10 [75]	D11 [4]	D12 [130]
Arabidopsis thaliana	0.716 (62)	0.958 (68)	0.803 (59)	—	—	—	—	—	1.464* (73)	—	0.404 (66)	—	1.270* (61)
Brassica napus	0.663 (62)	0.981 (66)	0.607 (61)	—	—	—	—	—	0.800 (66)	—	0.845 (62)	—	0.963 (61)
Sinapsis alba	0.730 (62)	0.992 (68)	0.705 (56)	—	—	—	—	—	1.127 (68)	—	0.845 (62)	—	0.855 (61)
Citrus limon	1.028 (70)	1.229 (75)	0.887 (61)	—	—	—	—	—	1.105* (76)	—	0.055 (69)	—	1.280* (72)
Fragaria ×ananassa	0.685 (66)	1.051 (72)	0.859 (56)	—	—	—	—	—	1.063* (72)	—	0.130 (64)	—	1.178 (66)
Lycopersicon esculentum	0.912 (68)	0.964 (73)	1.010 (58)	—	—	—	—	—	1.049* (75)	—	0.150 (64)	—	1.081 (67)
Oryza sativa	0.587 (71)	1.032 (79)	1.017 (64)	—	—	—	—	—	1.473† (82)	—	0.443 (68)	—	1.685† (80)

* RSF is not statistically significant after Markov correction for dinucleotide composition.
† Statistically significant RSF.

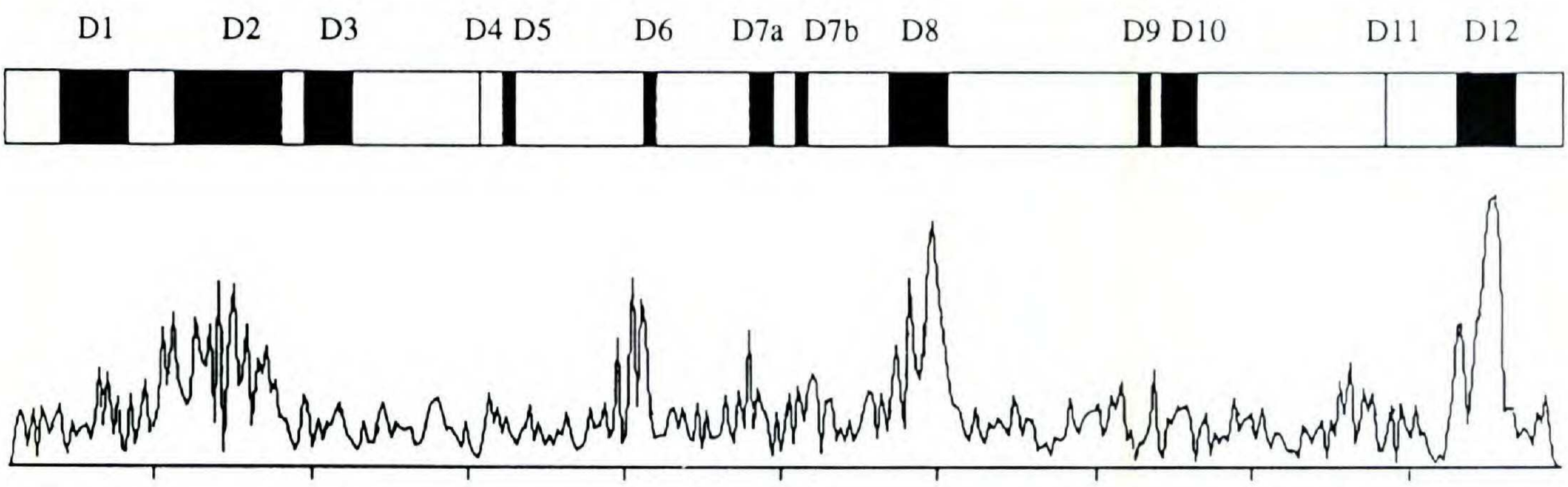


FIGURE 1. Locations of divergent domains (D1–D12) in the large subunit rDNA of *Oryza sativa* mapped to regions of sequence simplicity measured by SIMPLE34 (Hancock & Armstrong, 1994). Length of sequence given in base pairs.

involved 10 contiguous nucleotide positions in the alignment.

Overall sequence variation was much greater in the expansion segments than in the conserved core regions. Of 1034 nucleotide positions comprising the expansion segments, 431 (42%) were variable. Seventy of the 431 variable positions were due to insertion/deletion events. Of the 2423 nucleotides in the conserved core regions, 239 (10%) were variable. Forty-eight of the 239 variable positions in the conserved core regions were due to insertion/deletion events.

DOT PLOT ANALYSIS

A visual comparison of the dot plots of the LSU in plants to those for humans and bacteria reveals that the levels of internal sequence similarity are generally low in plants (Fig. 2). *Oryza* shows the strongest patterns of inter-expansion segment similarity, which appear as intensely dark regions on the dot plot. These regions of high similarity correspond well to the peaks of sequence simplicity associated with the expansion segments in the sequence simplicity profiles that border the dot plot. Interestingly, the dot plot of *Citrus limon* also shows a relatively strong pattern of expansion segment similarity. Because the RSF values for *Citrus* are not statistically significant, the similarity is most likely due to similar nucleotide composition in the expansion segment regions. This result highlights the importance of interpreting dot plot analyses in conjunction with sequence simplicity analyses.

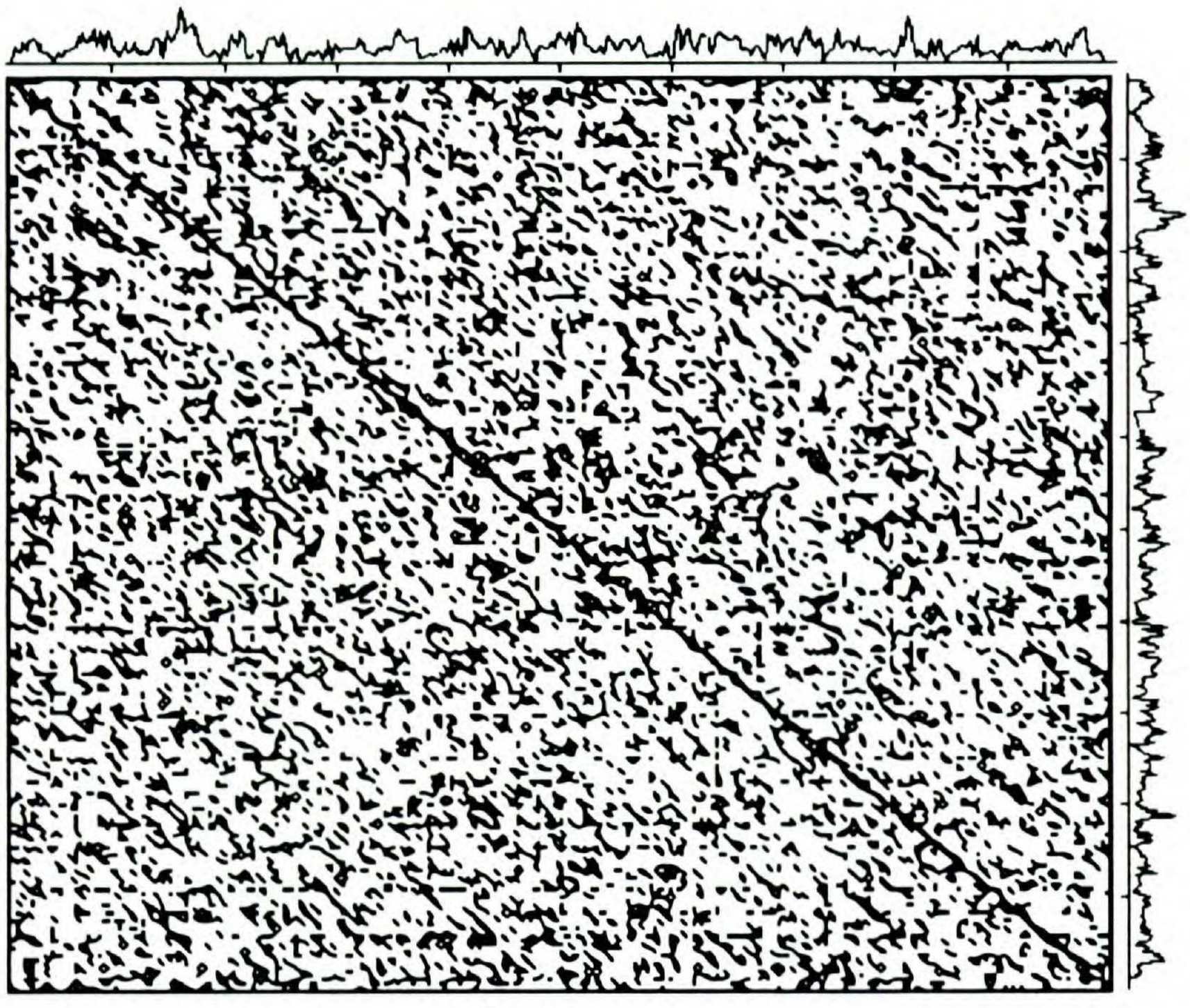
SUMMARY

The LSU of rDNA is often touted, anecdotally and in the literature, as an alternative gene to the small subunit (16/18S or SSU) of rDNA for phylogenetic analysis because it is longer and has a higher overall rate of sequence variability and, therefore, may provide more phylogenetically informative characters than the SSU for certain systematic questions (Larson, 1991). As demonstrated in this paper, and in the others cited herein, the patterns of nucleotide sequence change in the LSU are more complicated than is indicated by this "face value" assessment. Although the expansion segment regions of the LSU are potentially problematic as molecular markers in phylogenetic analyses, the conserved core regions may be useful as a source of molecular characters for exploring deep evolutionary divergences in the plant kingdom.

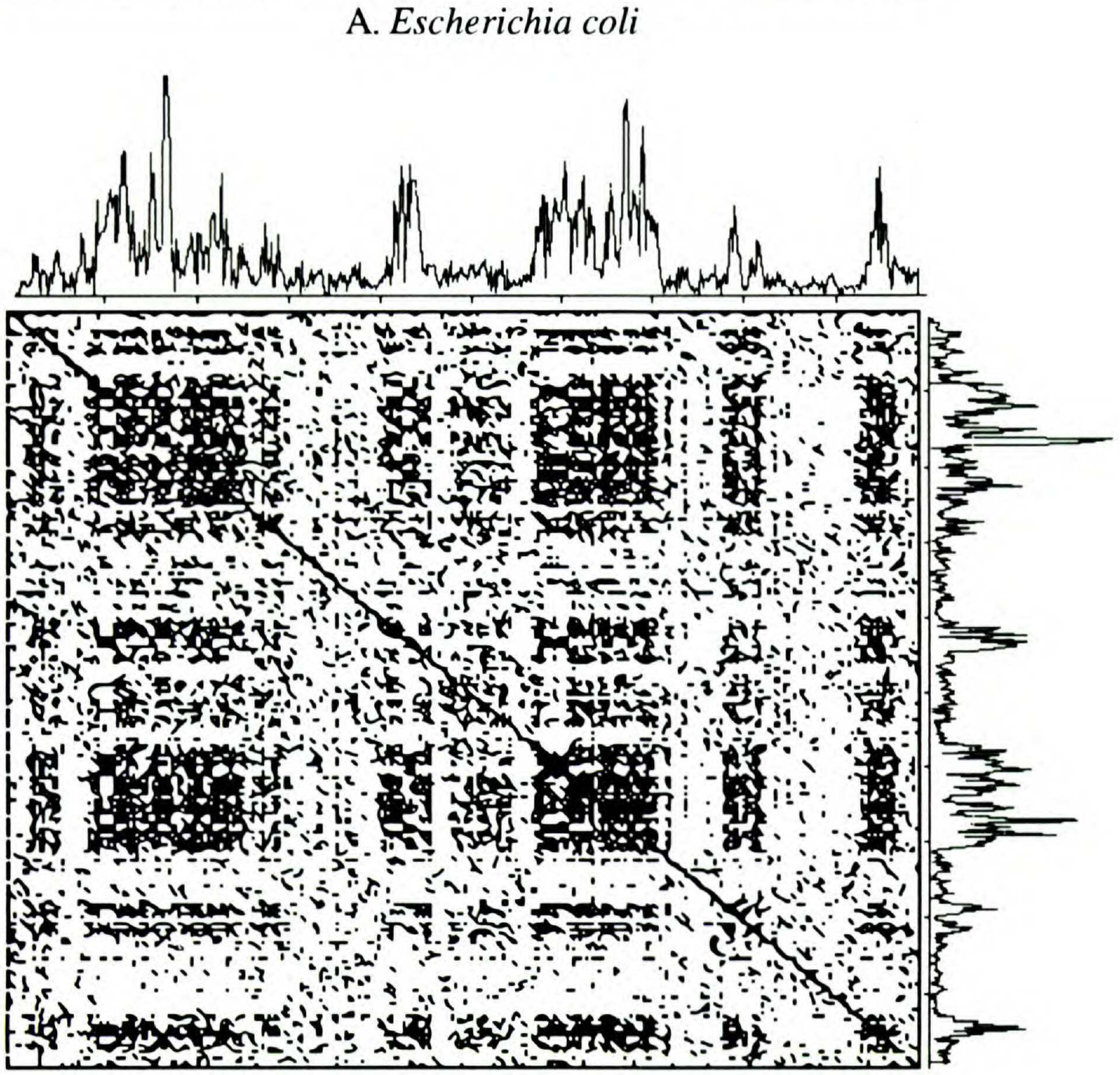
The analysis of the LSU for plants reveals the following: (1) nonrandom distribution and bias in nucleotide composition in expansion segments of all plant taxa examined, (2) significant cryptic sequence simplicity for *Oryza* and borderline simplicity for *Arabidopsis*, *Fragaria*, and *Sinapsis*, and (3) similarity among expansion segments (in *Oryza*) beyond that due to shared nucleotide composition. Nucleotide bias will complicate estimates of divergence times and will result in high levels of homoplasy due to convergence. For example, interspecific dot plot analyses by Hancock & Dover (1988, 1990) revealed a high degree of sequence similarity between human and rice due to nucle-

→

FIGURE 2A–I (pp. 240–244). Intra-specific dot plots of the complete sequences of LSU rDNA for seven plant species. Sequence simplicity profiles generated by SIMPLE34 border each dot plot. The higher the peak, the greater the level of sequence simplicity. Peaks of sequence simplicity generally correspond to expansion segments in eukaryotes. Dot plots for *Homo sapiens* and *Escherichia coli* are shown for comparison (see text).

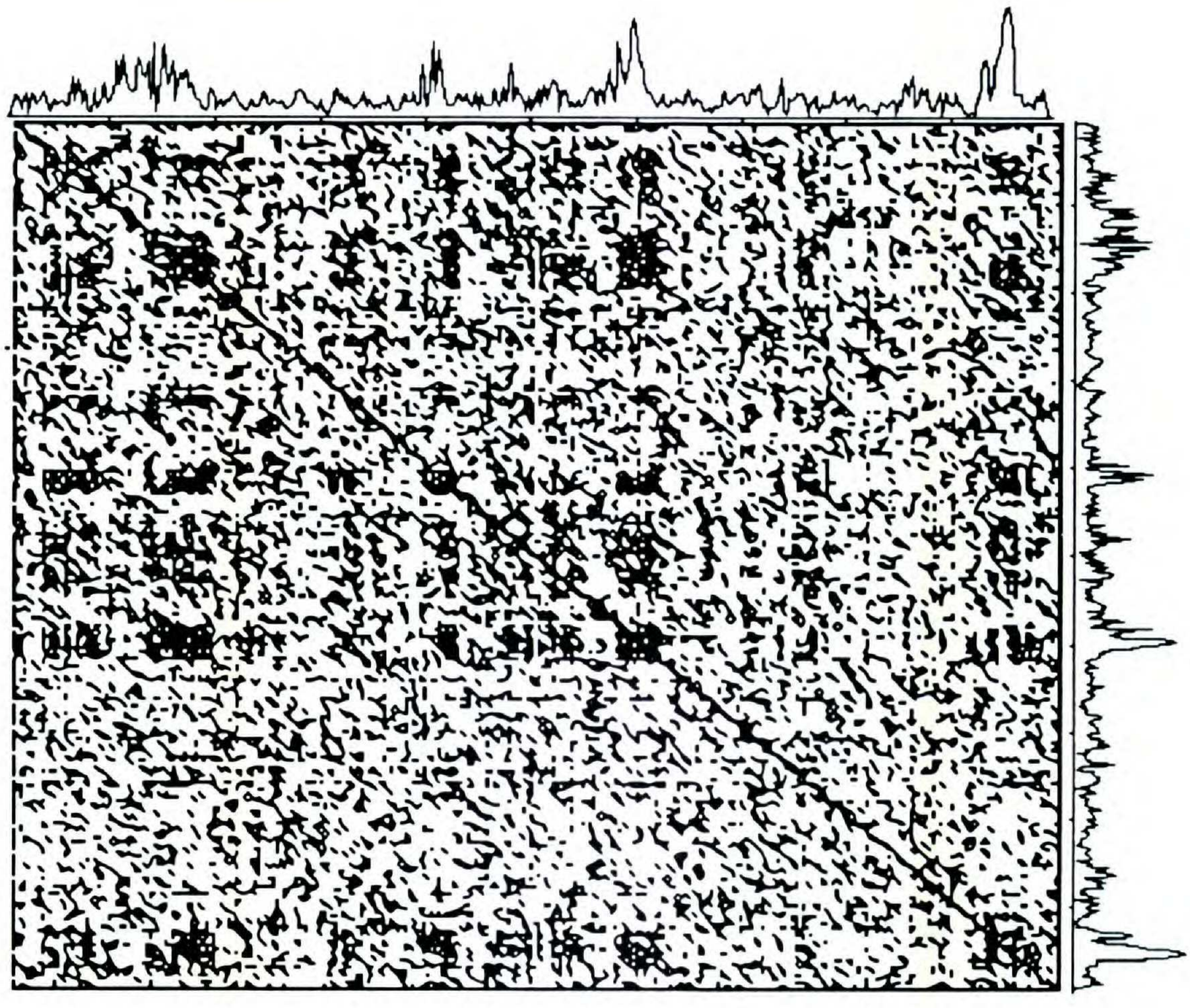


A. *Escherichia coli*

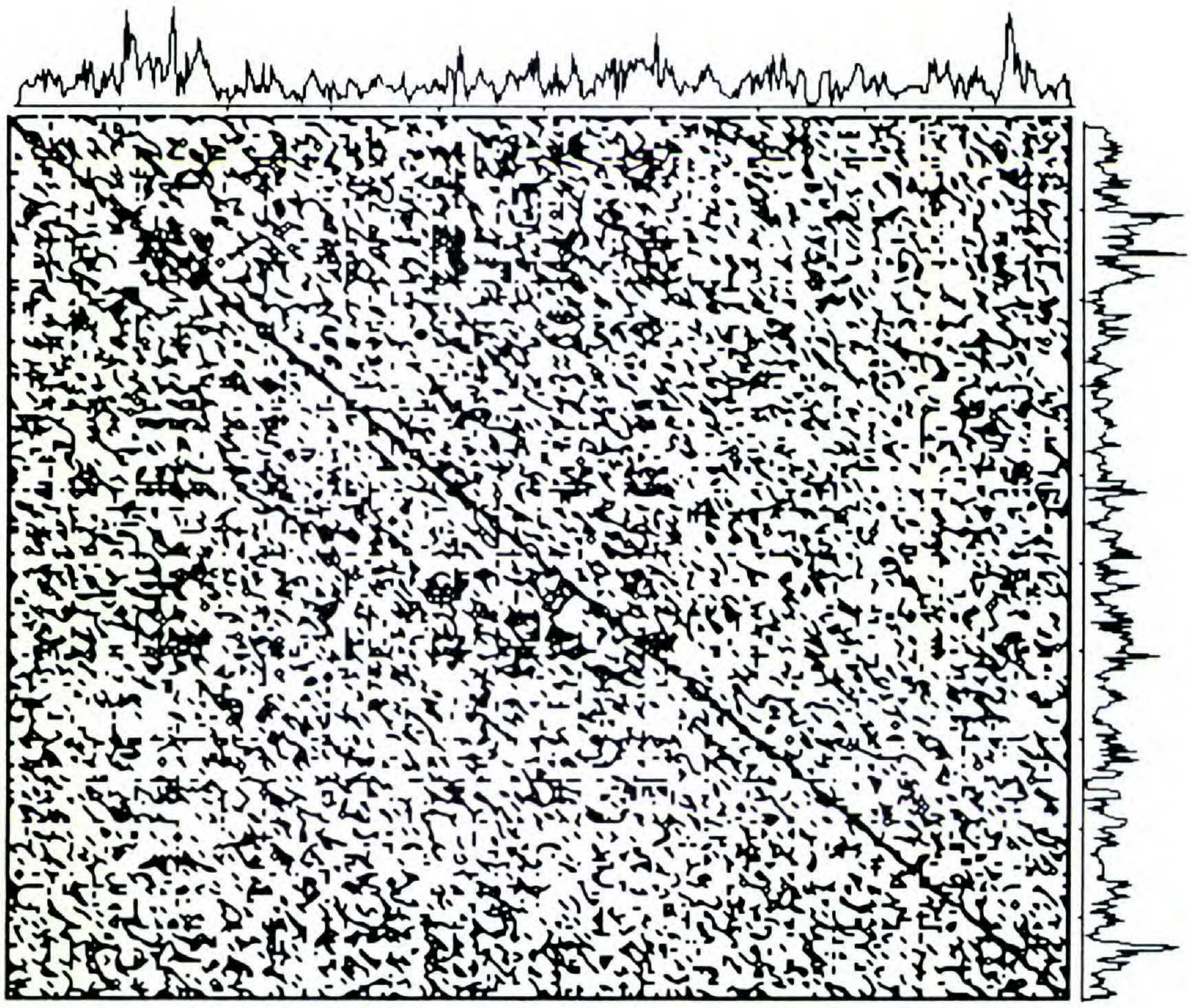


B. *Homo sapiens*

FIGURE 2. Continued.

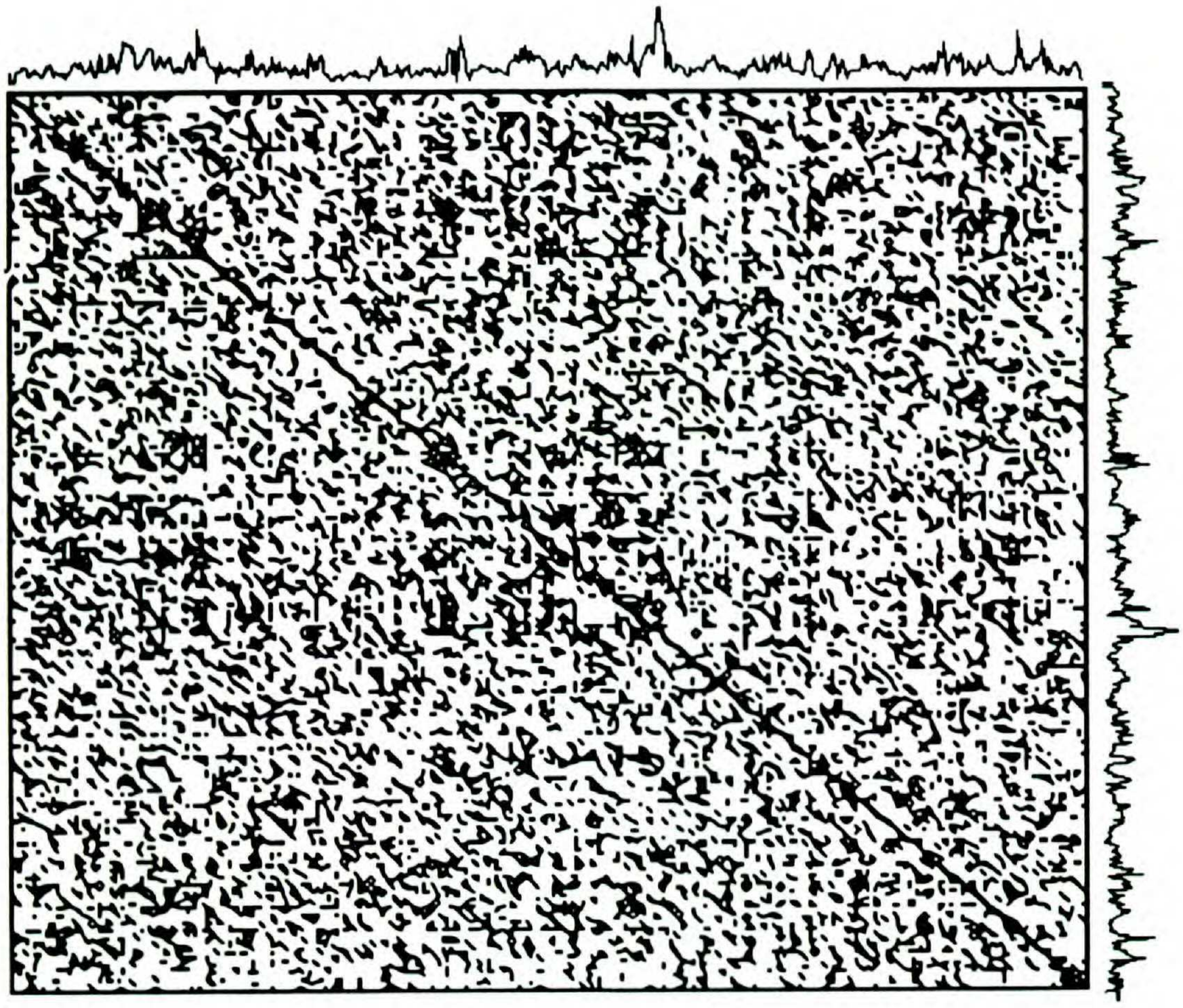


C. *Oryza sativa*



D. *Fragaria* X *ananassa*

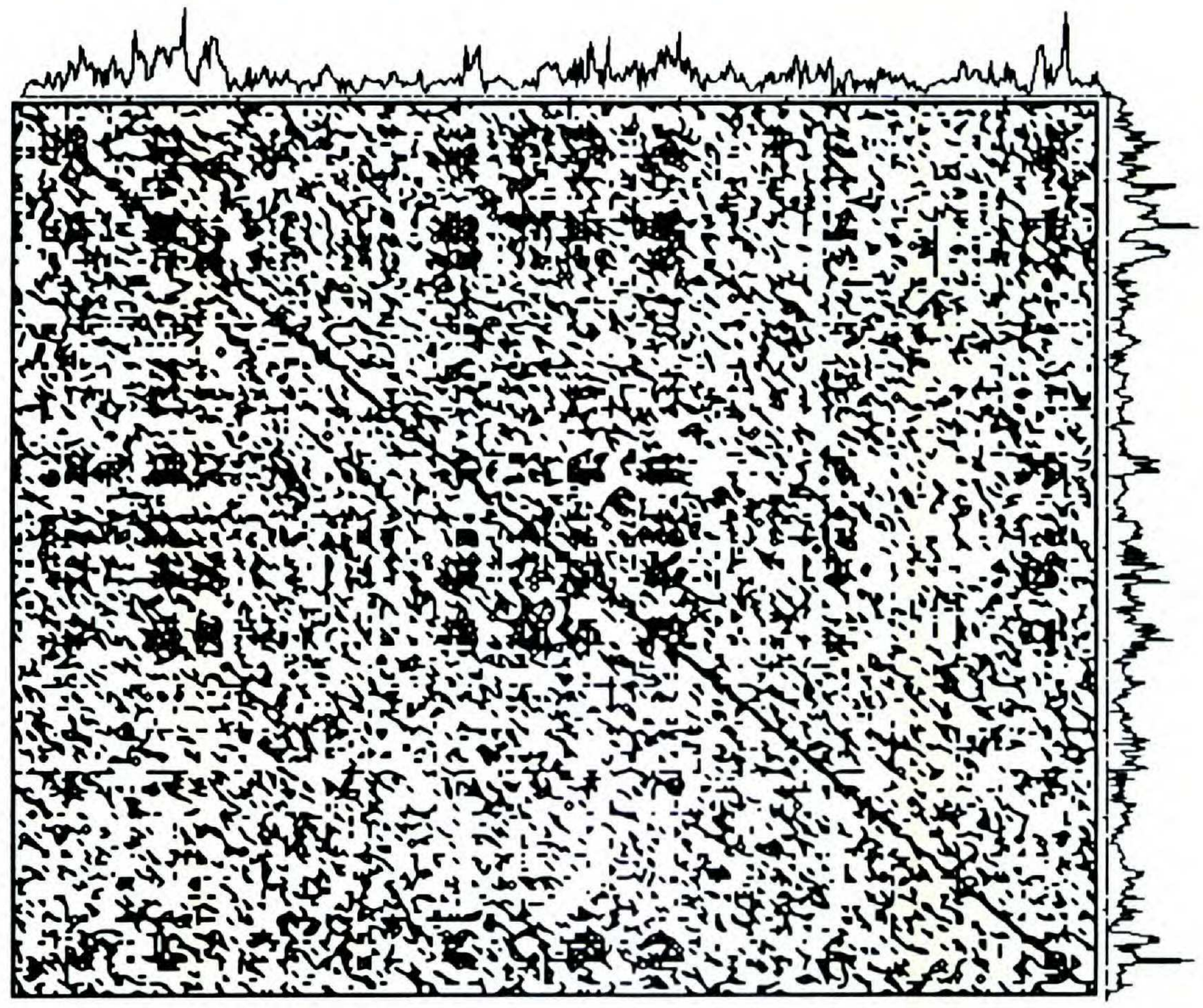
FIGURE 2. Continued.



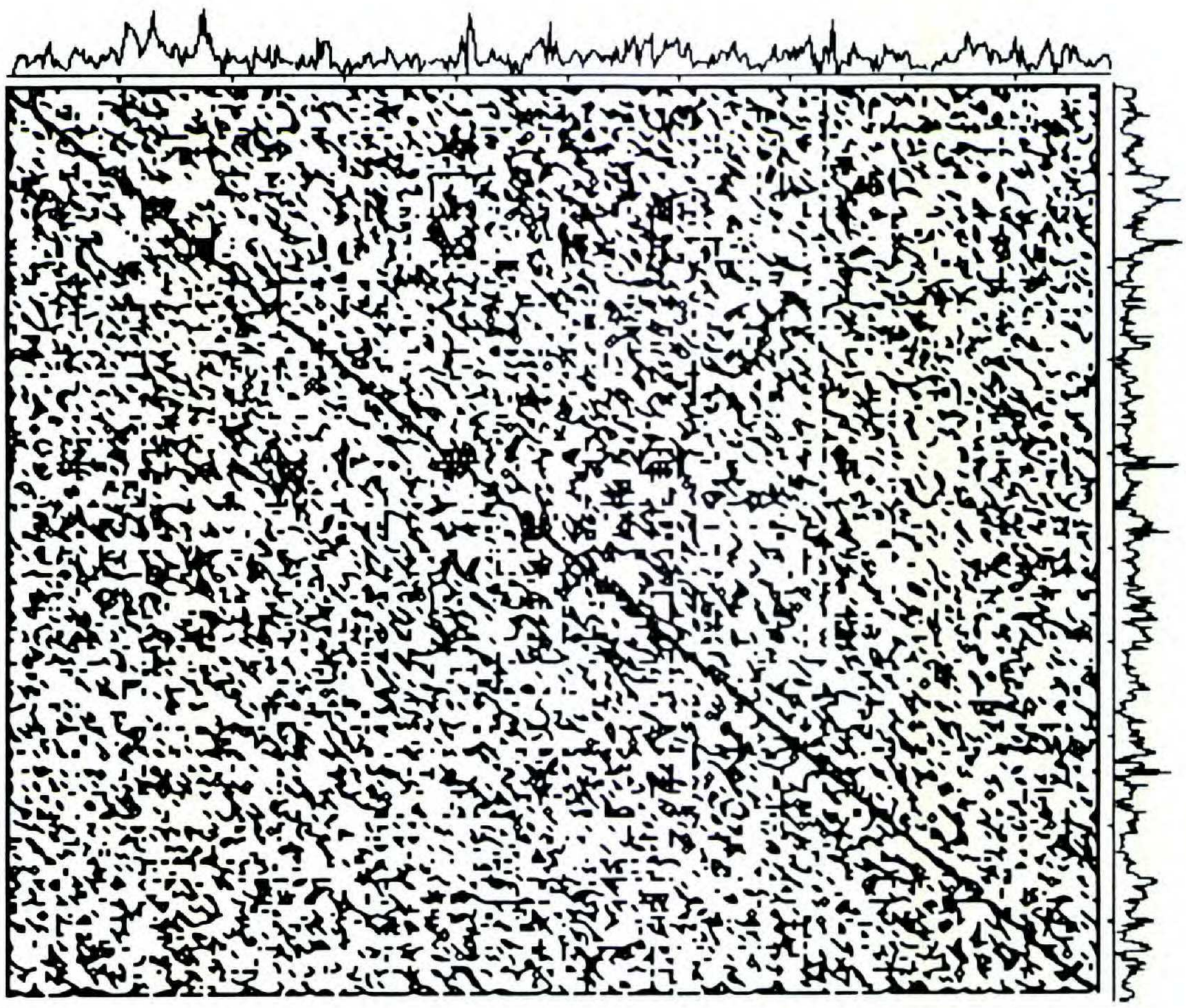
E. *Arabidopsis thaliana*

F. *Brassica napus*

FIGURE 2. Continued.

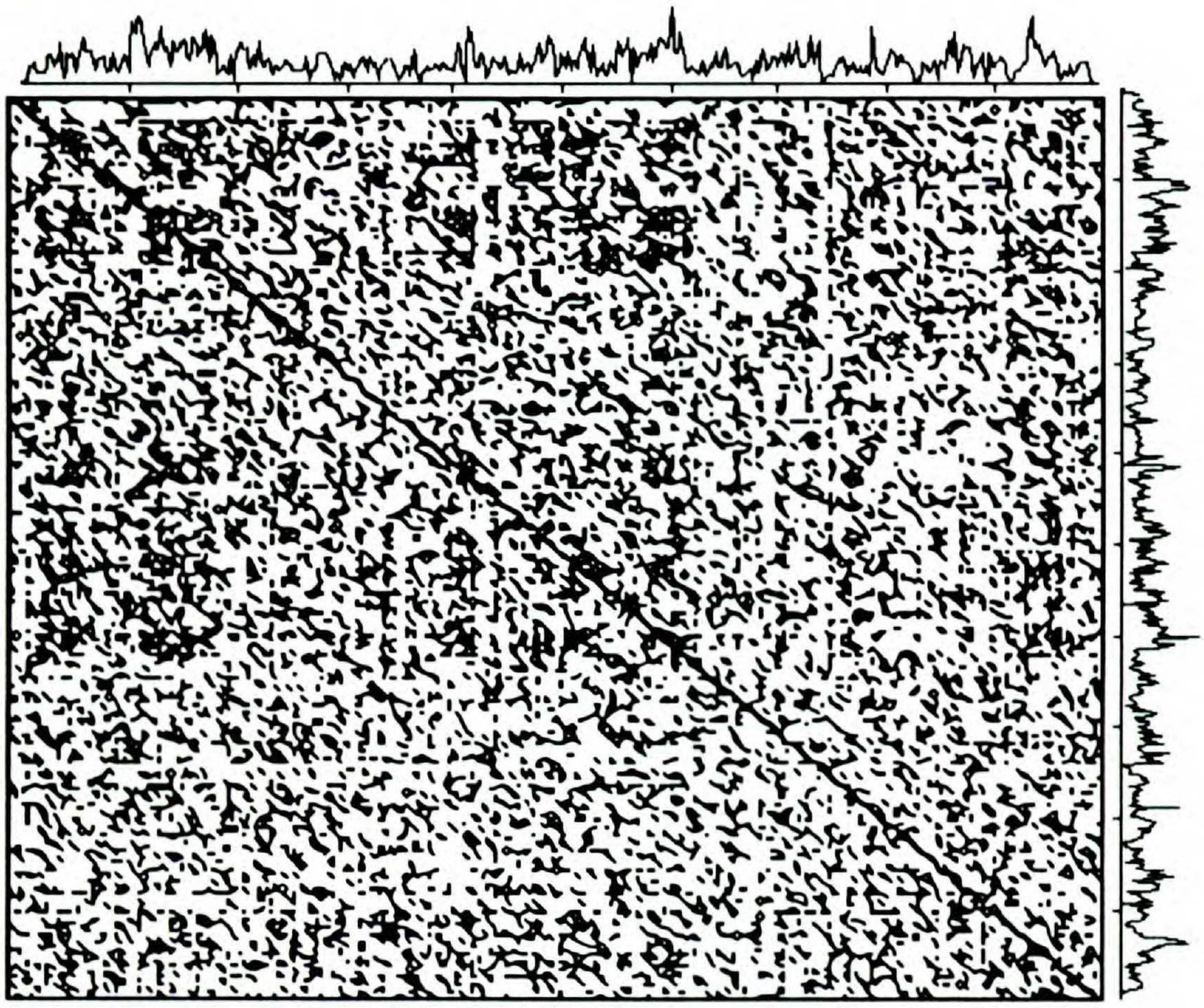


G. *Citrus limon*



H. *Sinapsis alba*

FIGURE 2. Continued.



I. *Lycopersicon esculentum*

FIGURE 2. Continued.

otide composition bias. Slippage-like mechanisms that lead to shuffling of small direct repeats in the expansion segments during turnover will obliterate evidence of ancestry in some cases, making homology assessment simply guesswork. This problem will be especially acute for deep divergences. Molecular coevolution among expansion segments will result in non-independence of characters. In their analysis of the secondary structure of the LSU in *Drosophila*, Hancock et al. (1988) described evidence for compensatory change among positions in the expansion segments. Weighting schemes have been proposed for compensatory mutations in the base-paired stem regions of rDNA based on secondary structure predictions (Wheeler & Honeycutt, 1988; Dixon & Hillis, 1993). Similar weighting schemes may be useful for expansion segments as well. The extent to which there is compensatory mutation in the expansion segments in plants is not known and awaits detailed modeling of secondary structure. In his molecular systematic work with the LSU of rDNA in salamanders, Larson (1991) explored the rates of sequence change in the expansion segments and concluded that violations to assumptions of parsimony were not severe enough to affect adversely the outcome of the analysis. He also proposed that a rate-invariant parsimony based on compositional statistics (Sidow & Wilson, 1990) be employed for phylogenetic analyses based on the LSU to compensate for the nucleotide composition bias observed in this molecule, especially for deep divergences. These post-alignment approaches, however, do not address the difficulties presented by the LSU for the basic assumptions of homology and independence among characters.

Because the number of plant species for which complete LSU sequence data currently are available is small and not representative of the overall phylogenetic diversity of the plant kingdom, we are sequencing the complete LSU of rDNA from representatives of all the major plant lineages. These data will enable us to determine if the patterns of sequence change described in this paper are found throughout the plant kingdom or are lineage-specific, and to test empirically the phylogenetic utility of both the expansion segments and the conserved core regions in the LSU of rDNA in plants.

LITERATURE CITED

APPELS, R. & R. L. HONEYCUTT. 1986. rDNA: Evolution over a billion years. Pp. 81–135 *in* S. K.

Dutta (editor), DNA Systematics, Vol. 2. CRC Press, New York.

ARNHEIM, N. 1983. Concerted evolution of multi-gene families. Pp. 38–61 *in* M. Nei & R. K. Koehn (editors), Evolution of Genes and Proteins. Sinauer Associates, Sunderland, Massachusetts.

BALDWIN, B. G. 1992. Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: An example from Compositae. Molec. Phylogenetics Evol. 1: 3–16.

———, M. J. SANDERSON, J. M. PORTER, M. F. WOJCIECHOWSKI, C. S. CAMPBELL & M. J. DONOGHUE. 1995. The ITS region of nuclear ribosomal DNA: A valuable source of evidence on angiosperm phylogeny. Ann. Missouri Bot. Gard. 82: 247–277.

BROSIUS, J., T. J. DULL & H. F. NOLLER. 1980. Complete nucleotide sequence of a 23S ribosomal RNA gene from *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A. 77: 201–204.

BULT, C. J. & E. A. ZIMMER. 1993. Nuclear ribosomal RNA sequences for inferring tribal relationships within Onagraceae. Syst. Bot. 18: 48–63.

CLARK, G., B. W. TAGUE, V. C. WARE & S. A. GERBI. 1984. *Xenopus laevis* 28S ribosomal RNA: A secondary structure and its evolutionary and functional implications. Nucl. Acids Res. 12: 6197–6220.

DIXON, M. T. & D. M. HILLIS. 1993. Ribosomal RNA secondary structure: Compensatory mutations and implications for phylogenetic analysis. Molec. Biol. Evol. 10: 256–267.

DOVER, G. A. & R. B. FLAVELL. 1984. Molecular coevolution: DNA divergence and the maintenance of function. Cell 38: 622–623.

FLAVELL, R. B. 1986. Structure and control of expression of ribosomal RNA genes. Oxford Surv. Pl. Molec. Cell. Biol. 3: 252–274.

FRIEDLANDER, T. P., J. C. REGIER & C. MITTER. 1992. Nuclear gene sequences for higher level phylogenetic analysis: 14 promising candidates. Syst. Biol. 41: 483–490.

GERBI, S. A. 1985. Evolution of ribosomal DNA. Pp. 419–517 *in* R. J. MacIntyre (editor), Molecular Evolutionary Genetics. Plenum Press, New York.

GONZALEZ, I. L., J. L. GORSKI, T. J. CAMPEN, D. J. DORNEY, J. M. ERICKSON, J. E. SYLVESTER & R. D. SCHMICKEL. 1985. Variation among human 28S ribosomal RNA genes. Proc. Natl. Acad. Sci. U.S.A. 82: 7666–7670.

HAMBY, R. K. & E. A. ZIMMER. 1988. Ribosomal RNA sequences for inferring phylogeny within the grass family (Poaceae). Pl. Syst. Evol. 160: 29–37.

HANCOCK, J. M. & J. S. ARMSTRONG. 1994. SIMPLE34: An improved and enhanced implementation for VAX and SUN computers of the SIMPLE algorithm for the analysis of clustered repetitive sequences. CABIOS 10: 67–70.

——— & G. A. DOVER. 1988. Molecular coevolution among cryptically simple expansion segments of eukaryotic 26S/28S rRNAs. Molec. Biol. Evol. 5: 377–391.

——— & ———. 1990. 'Compensatory slippage' in the evolution of ribosomal RNA genes. Nucl. Acids Res. 18: 5949–5954.

———, D. TAUTZ & G. A. DOVER. 1988. Evolution of the secondary structure and compensatory mutations of the ribosomal RNAs of *Drosophila melanogaster*. Molec. Biol. Evol. 5: 393–414.

HASSOUNA, N., B. MICHOT & J. P. BACHELLERIE. 1984. The complete nucleotide sequence of mouse 28S rRNA gene: Implications for the process of size increase of the large subunit rRNA in higher eukaryotes. Nucl. Acids Res. 12: 3563–3574.

HILLIS, D. M. & M. T. DIXON. 1991. Ribosomal DNA: Molecular evolution and phylogenetic inference. Quart. Rev. Biol. 66: 411–453.

JORGENSEN, R. A. & P. D. CLUSTER. 1988. Modes and tempos in the evolution of nuclear ribosomal DNA: New characters for evolutionary studies and new markers for genetic and population studies. Ann. Missouri Bot. Gard. 75: 1238–1247.

KANTZ, T., E. THERIOT, E. ZIMMER & R. L. CHAPMAN. 1990. The Pleurastrophyceae and Micromonadophyceae: A cladistic analysis of nuclear rRNA sequence data. J. Phycol. 26: 711–721.

KISS, T., M. KISS & F. SOLYMOSY. 1989. Nucleotide sequence of a 25S rRNA gene from tomato. Nucl. Acids Res. 17: 796.

KNAAK, C., R. K. HAMBY, M. L. ARNOLD, M. D. LEBLANC, R. L. CHAPMAN & E. A. ZIMMER. 1990. Ribosomal DNA variation and its use in plant biosystematics. Pp. 135–158 *in* S. Kawano (editor), Biological Approaches and Evolutionary Trends in Plants. Academic Press, New York.

KOLOSHA, V. O. & I. I. FODOR. 1990. Nucleotide sequence of *Citrus limon* 26S rRNA gene and secondary structure model of its RNA. Pl. Molec. Biol. 14: 147–161.

LARSON, A. 1991. Evolutionary analysis of length-variable sequences: Divergent domains of ribosomal RNA. Pp. 221–247 *in* M. Miyamoto & J. Craycraft (editors), Phylogenetic Analysis of DNA Sequences. Oxford Univ. Press, New York.

NICKRENT, D. L. & D. E. SOLTIS. 1995. A comparison of angiosperm phylogenies from nuclear 18S rDNA and *rbc*L sequences. Ann. Missouri Bot. Gard. 82: 208–234.

OKUMURA, S. & H. SHIMADA. 1992. Nucleotide sequence of genes for 5.8S and 25S rRNA from rapeseed (*Brassica napus*). Nucl. Acids Res. 20: 3510.

SCHAAL, B. A. & G. H. LEARN, JR. 1988. Ribosomal DNA variation within and among plant populations. Ann. Missouri Bot. Gard. 75: 1207–1216.

SIDOW, A. & A. C. WILSON. 1990. Compositional statistics: An improvement in evolutionary parsimony and its application to deep branches in the tree of life. J. Molec. Evol. 31: 51–68.

SUGIURA, M., Y. IIDA, K. OONO & F. TAKAIWA. 1985. The complete nucleotide sequence of a rice 25S rRNA gene. Gene 37: 255–259.

TAUTZ, D., M. TRICK & G. A. DOVER. 1986. Cryptic simplicity in DNA is a major source of genetic variation. Nature 322: 652–656.

———, J. M. HANCOCK, D. A. WEBB, C. TAUTZ & G. A. DOVER. 1988. Complete sequences of the rRNA genes of *Drosophila melanogaster*. Molec. Biol. Evol. 5: 366–376.

UNFRIED, I. & P. GRUENDLER. 1990. Nucleotide sequence of the 5.8S and 25S rRNA genes and of the internal transcribed spacers from *Arabidopsis thaliana*. Nucl. Acids Res. 18: 4011.

WHEELER, W. C. & R. L. HONEYCUTT. 1988. Paired sequence difference in ribosomal RNAs: Evolutionary and phylogenetic implications. Molec. Biol. Evol. 5: 90–96.

ZECHMAN, F., E. THERIOT, E. ZIMMER & R. CHAPMAN. 1990. Phylogeny of the Ulvophyceae (Chlorophyta): Cladistic analysis of nuclear-encoded rRNA sequence data. J. Phycol. 26: 700–710.

ZIMMER, E., R. HAMBY, M. ARNOLD, D. LEBLANC & E. THERIOT. 1989. Ribosomal RNA phylogenies and flowering plant evolution. Pp. 205–226 *in* B. Fernholm, K. Bremer & H. Jornvall (editors), The Hierarchy of Life: Molecules and Morphology in Phylogenetic Analysis. Excerpta Medica, New York.